Lesson 3.2.3

# Investigating the Meaning of Numbers in the Equation of a Line

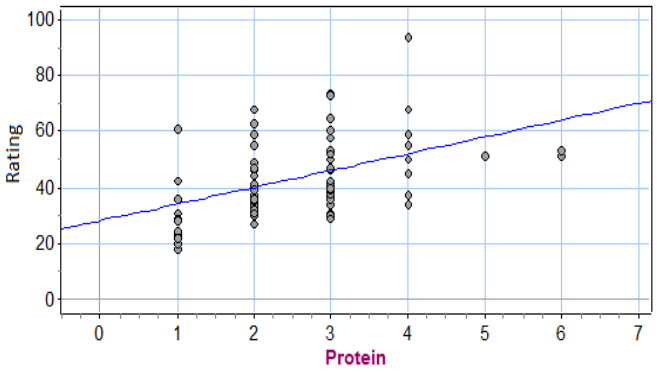
STUDENT NAME _____ DATE _____

## TRY THESE

### Investigating the Line

1 Here you return to the cereal data. Begin with a few questions to review what you know about the least squares line. Here is the least squares line that predicts *Consumer Reports* ratings based on the amount of protein (in grams) in a serving:

- Predicted rating = 28 + 6(protein).
- The correlation coefficient is 0.48.




A Use the least squares regression line to predict the rating for a cereal containing 2 grams of protein in a serving.

B There are two cereals with 6 grams of protein in a serving. Is the predicted rating from the least squares regression line too high or too low for these cereals?

C What does the phrase *least squares* tell you about this line?

**© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING**

Lesson 3.2.3
## Investigating the Meaning of Numbers in the Equation of a Line

D Which of these statements would you use to describe protein as a predictor?

- Protein is a very accurate predictor of *Consumer Reports* ratings (errors within a few rating points would be typical).
- Protein is not a very accurate predictor of ratings (errors as large as 10 rating points would not be surprising)

2 Now you will focus on understanding what information you get from the numbers in the equation of the line. Here we have the equation of the least squares line and a table of protein amounts and predicted ratings from the least squares line.

predicted rating = 28 + 6(protein)

$$\hat{y} = 28 + 6x$$

| $x$ = protein (g/serving) | $\hat{y}$= predicted rating |
|---|---|
| 0 | 28 |
| 1 | 34 |
| 2 | 40 |
| 3 | 46 |
| 4 | 52 |
| 5 | 58 |

How are the 28 and 6 related to the table of values? Be as specific as you can.

3 Here you have the least squares equation for predicting *Consumer Reports* ratings based on the amount of sugar (in grams) in a serving. The equation was used to generate the table of predicted values for some sugar amounts.

predicted rating = 60 – 2.4(sugars)

$$\hat{y} = 60 - 2.4x$$

| $x$ = sugar (g/serving) | $\hat{y}$= predicted rating |
|---|---|
| 0 | 60 |
| 1 | 57.6 |
| 2 | 55.2 |
| 3 | 52.8 |
| 4 | 50.4 |
| 5 | 48 |

How are the 60 and –2.4 related to the table of values? Be as specific as you can.

In the next set of tasks, you will continue to focus on understanding the numbers in the equation of a line.

**© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING**

Lesson 3.2.3
# Investigating the Meaning of Numbers in the Equation of a Line

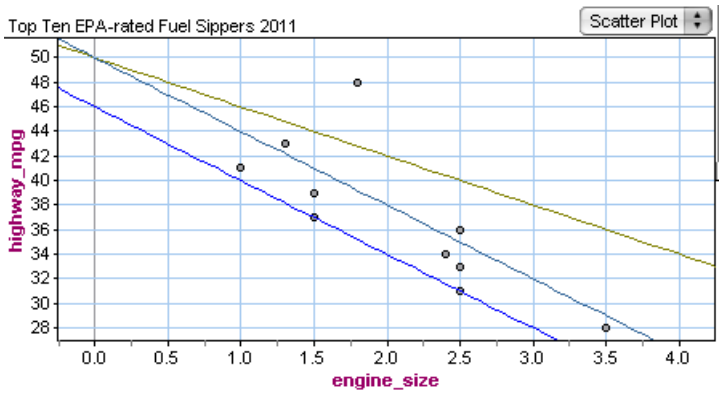
## NEXT STEPS

### Identifying the Best Fit Line

4 The Environmental Protection Agency picks the 10 most fuel-efficient cars each year. Below is a scatterplot of the highway miles per gallon and the engine size (measured in liters) for the EPA's top 10 for 2011. (Retrieved from www.fueleconomy.gov)

Following are the equations of the three lines shown:

$$\hat{y} = 46 - 6x$$
$$\hat{y} = 50 - 6x$$
$$\hat{y} = 50 - 4x$$




A Identify the *equation* of the line that best fits the data. Briefly explain how you made your decision.

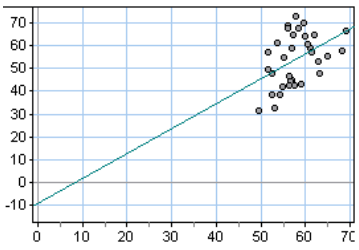
B For the line you chose, describe what the numbers in the equation tells you about engine size, highway miles per gallon, and the relationship between the two for these 10 cars.

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

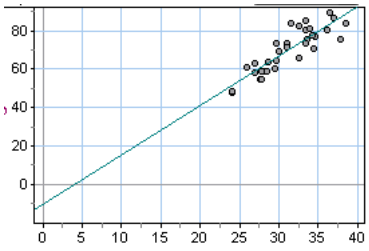
Lesson 3.2.3

# Investigating the Meaning of Numbers in the Equation of a Line

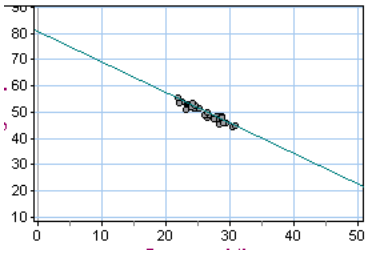
5 Match the graphs to the least squares equations and *r*-values.



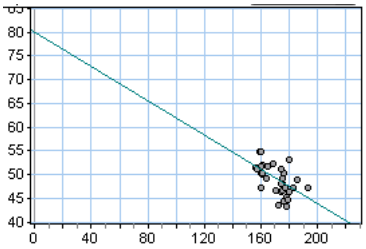
**Graph A**



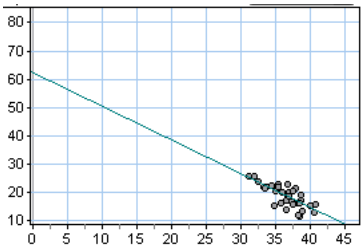
**Graph B**



**Graph C**



**Graph D**



**Graph E**

Here are $r$-values to choose from:

| | –0.54 | –0.73 | 0.45 | –0.95 | 0.88 |
|---|---|---|---|---|---|
| **Graph:** | | | | | |

Here are equations to choose from:

| | $\hat{y} = -10.5 + 2x$ | $\hat{y} = 62 + (-1.2)x$ | $\hat{y} = -10.5 + 1.1x$ | $\hat{y} = 80 + (-1.2)x$ | $\hat{y} = 80 + (-0.2)x$ |
|---|---|---|---|---|---|
| **Graph:** | | | | | |

**© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING**

Lesson 3.2.3

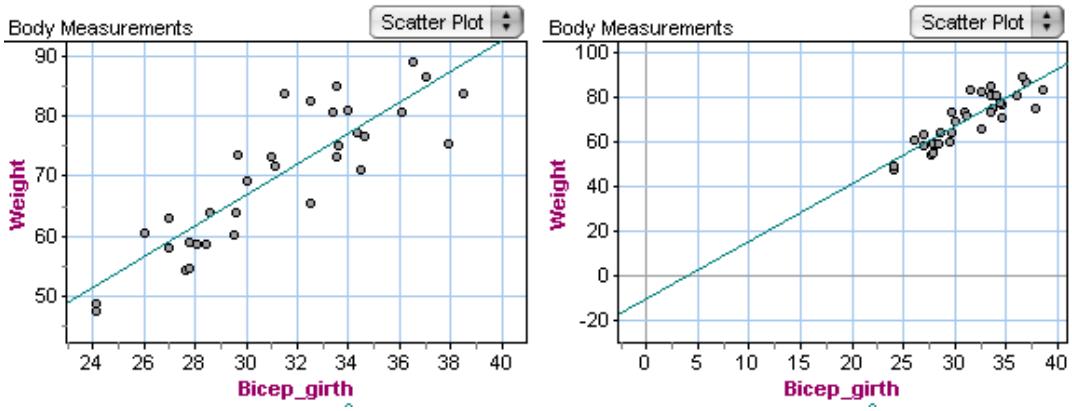
# Investigating the Meaning of Numbers in the Equation of a Line

## TAKE IT HOME

1 Based on data from 34 adults who exercise regularly, the least squares line for the relationship between bicep girth and weight is

$$\text{predicted weight} = 2.6(\text{bicep girth}) - 10.5$$

where predicted weight is measured in kilograms and bicep girth is measured in centimeters.




A Construct a table or use one of the graphs to explain the meaning of 2.6 in this situation.

B In the prediction equation, –10.5 is the initial value when $x = 0$. Does this number have meaning in this scenario? Why or why not?

C If you use bicep girth to predict weight, how accurate do you think the predictions will be?

- Very accurate (typical prediction error will be within a kilogram).
- Somewhat accurate (typical prediction error will be within 10 kilograms).
- Not very accurate (prediction errors larger than plus or minus 20 kilograms would not be surprising).

© 2011 THE CARNEGIE FOUNDATION FOR THE ADVANCEMENT OF TEACHING

Lesson 3.2.3
**Investigating the Meaning of Numbers in the Equation of a Line**

2 With the following applet, you will investigate how outliers impact the regression line.

http://www.stat.sc.edu/~west/javahtml/Regression.html

A Add points to the scatterplot that are close data points shown. Describe what happens to the regression line.

B A data point is *influential* if removing it (or in this case adding it) substantially changes the regression line. Add points to the scatterplot that are outliers relative to the other data points. In other words, add points that are far away from the other data points. Describe what happens to the regression line.

+++++

This lesson is part of STATWAY™, A Pathway Through College Statistics, which is a product of a Carnegie Networked Improvement Community that seeks to advance student success. Version 1.0, A Pathway Through Statistics, Statway™ was created by the Charles A. Dana Center at the University of Texas at Austin under sponsorship of the Carnegie Foundation for the Advancement of Teaching. This version 1.5 and all subsequent versions, result from the continuous improvement efforts of the Carnegie Networked Improvement Community. The network brings together community college faculty and staff, designers, researchers and developers. It is an open-resource research and development community that seeks to harvest the wisdom of its diverse participants in systematic and disciplined inquiries to improve developmental mathematics instruction. For more information on the Statway Networked Improvement Community, please visit **carnegiefoundation.org.** For the most recent version of instructional materials, visit **Statway.org/kernel.**

+++++

STATWAY™ and the Carnegie Foundation logo are trademarks of the Carnegie Foundation for the Advancement of Teaching. A Pathway Through College Statistics may be used as provided in the CC BY

Lesson 3.2.3
**Investigating the Meaning of Numbers in the Equation of a Line**

license, but neither the Statway trademark nor the Carnegie Foundation logo may be used without the prior written consent of the Carnegie Foundation.